Space Data

Fernando Nascimento

2023-02-01

OUTLINE

- Executive Summary
- Introduction
- Methodology
- Results
- Discussion
- Conclusion
- Appendix

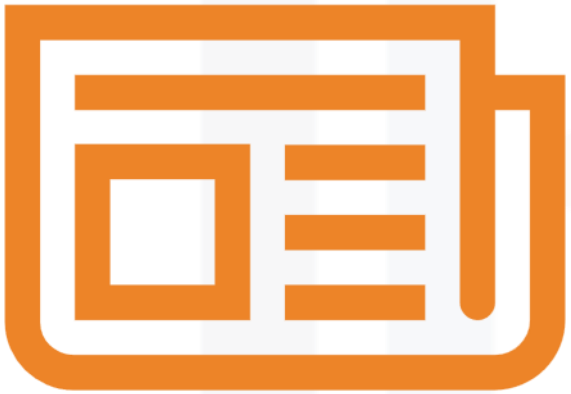
EXECUTIVE SUMMARY

- The data was analyzed by using SpaceX API and by Web Scraping.
- The following manipulation methods were used on the data:
  - Exploratory Data Analysis (EDA);
  - Data visualization - Including an interactive visual analysis;
  - Machine Learning Prediction.
- Explaining the methodologies:
  - EDA allows to find the best features to predict success launchings;
  - Visualize the data is the better way to show correlations between important topics;
  - Machine Learning predictions shows the most important characteristics to drive this opportunity by the best way.

INTRODUCTION

- The objective of this project is to evaluate and maximize the viability of the company to compete with the Space X
- Desirable answers:
  - Total cost of the Launches;
  - Best places to launch future spaceships;
  - Correct predictions of successful landings.

METHODOLOGY



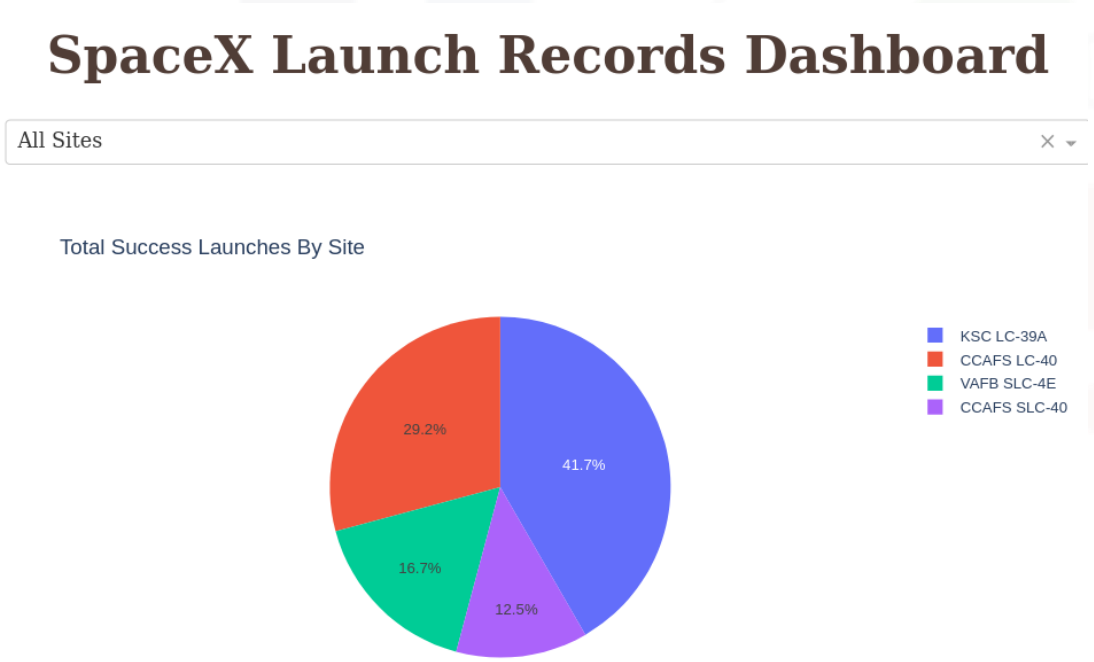
- Data Collection:
  - The data was obtained from 2 sources:
    - Space X API (https://api.spacexdata.com/v4/rockets/)
    - Web Scraping (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches)
- Data Wrangling
  - Collected data was enriched by creating a landing outcome label based on outcome data after summarizing and analyzing features
- Predictive Analysis
  - The Data was divided in training and test data sets and evaluated by four different classification models, being the accuracy of each model evaluated using different combinations of parameters.

RESULTS

- Exploratory data analysis results:
  - Space X uses 4 different launch sites;
  - The first launches were done to Space X itself and NASA;
  - The average payload of F9 v1.1 booster is 2,928 kg;
  - The first success landing outcome happened in 2015 fiver year after the first launch;
  - Many Falcon 9 booster versions were successful at landing in drone ships having payload above the average;
  - Almost 100% of mission outcomes were successful;
  - Two booster versions failed at landing in drone ships in 2015: F9 v1.1 B1012 and F9 v1.1 B1015;
  - The number of landing outcomes became as better as years passed.

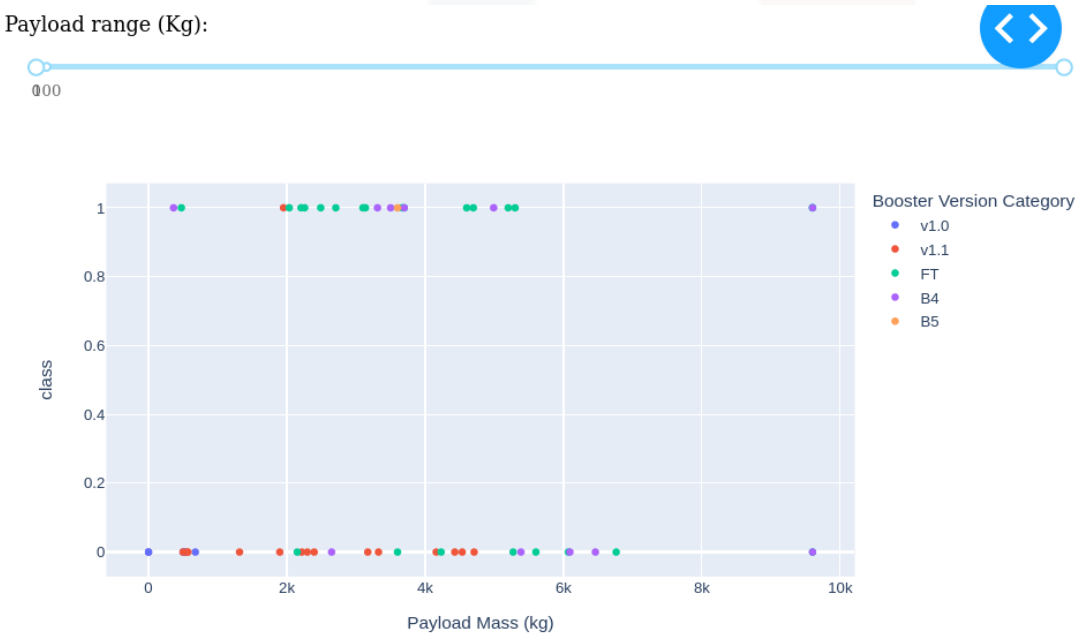
- Using interactive analytics was possible to identify that launch sites use to be in safety places, near sea, for example and have a good logistic infrastructure around.
- • Most launches happens at east cost launch sites.
- Predictive Analysis showed that Decision Tree Classifier is the best model to predict successful landings, having accuracy over 87% and accuracy for test data over 94%.

DASHBOARD



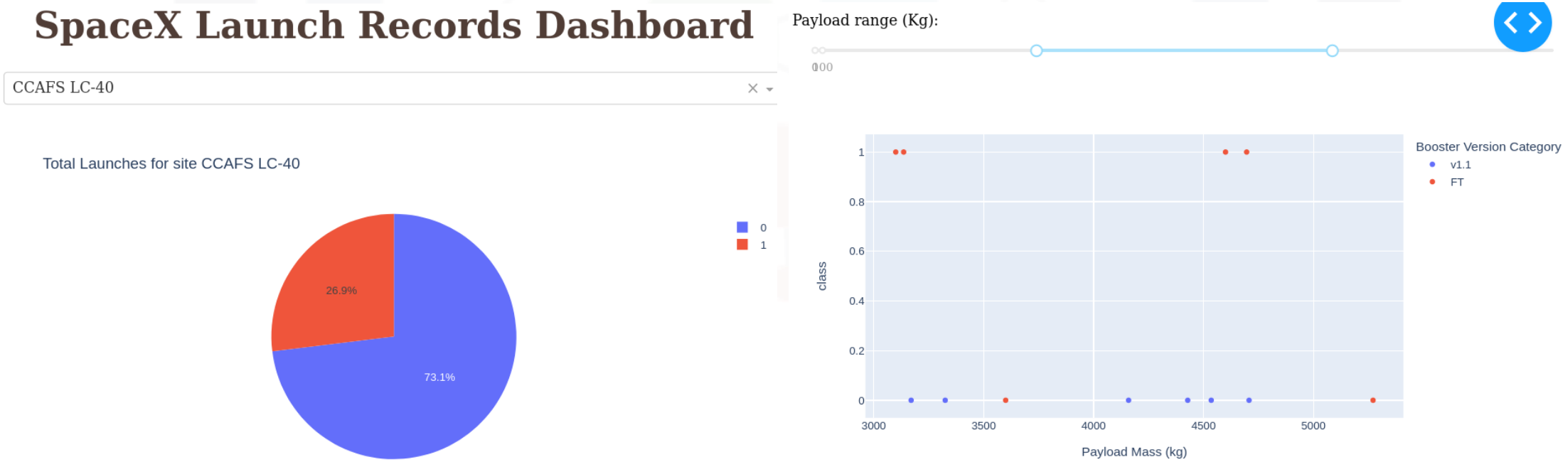
- The dashboard is an example on how we can visualize the data.
- The dropdown gives access to a specific Launch Site and the success rate of them individually.

DASHBOARD



- This interactive dashboard allows to "travel" the data by given a Payload range.
- The data in the dashboard is divided by the Booster Version, a subtitle is right next to the graph.

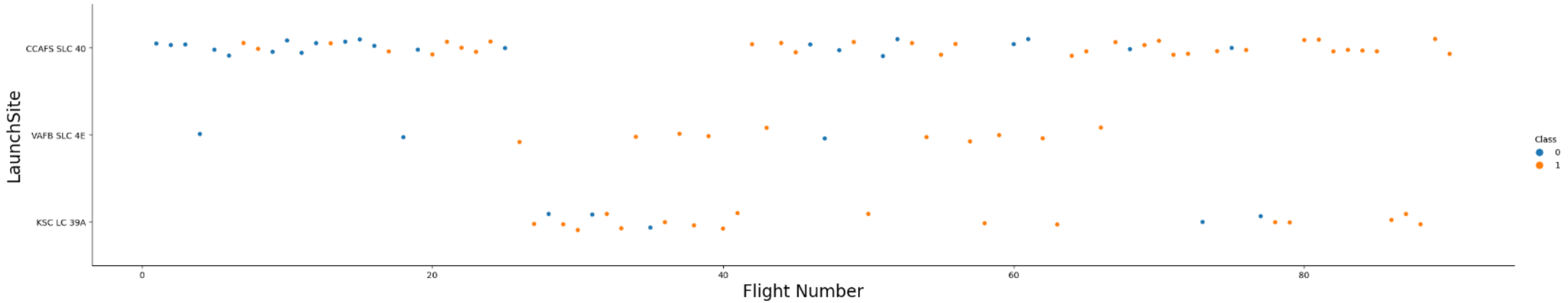
DASHBOARD - Examples



DISCUSSION

- The main purpose of the project is to see the correlation between the variables to evaluate and select the ones with the best success rate.

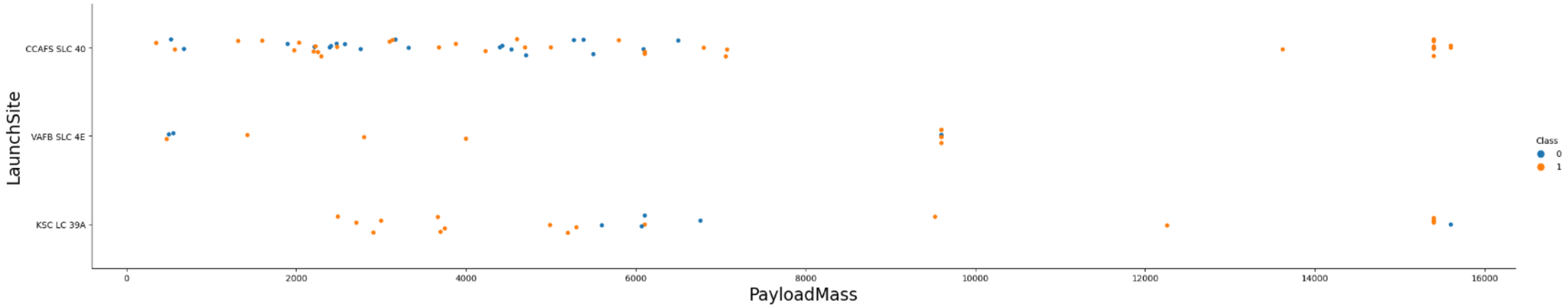
Launch Site x Flight Number



The plot above shows the best launch site and the general success rate progress as the time passes.

The best launch site is CCAFS SLC 40, followed by VAFB SLC 4E and KSC LC 39A.

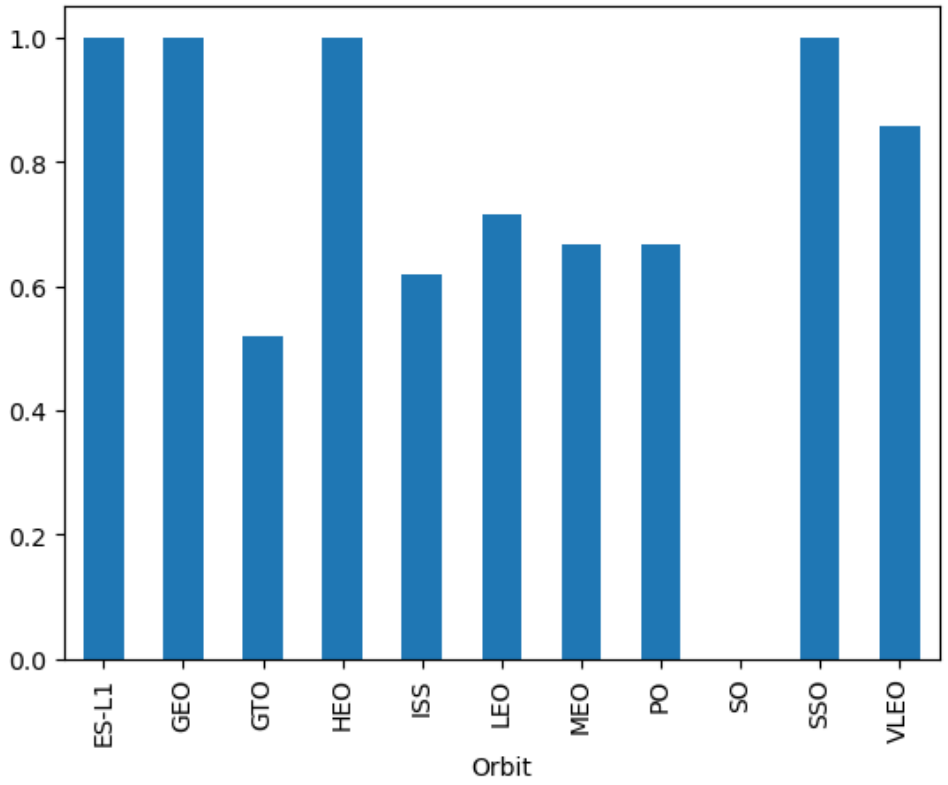
Launch Site x Payload Mass



The plot above shows that the average payload mass is below 8000 Kg and payloads above 12000 Kg have a very positive success rate.

Payloads above 12000 Kg seems to be not possible at VAFB SLC 4E launch site.

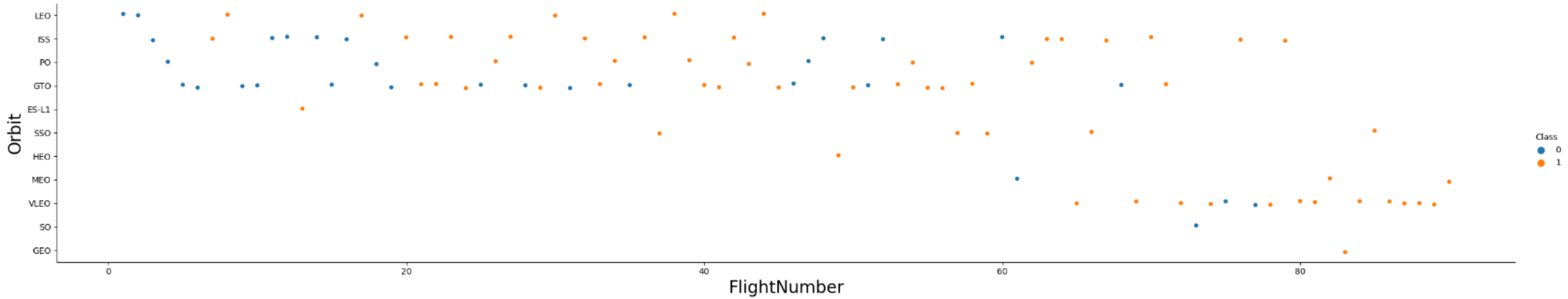
Success Rate x Orbit



The best success rate happens in the orbits: ES-11, GEO, HEO and SSO. All the orbits mentioned shows a 1/1 success rate.

The worst case it's seen at orbit SO that showed a 0/1 success rate.

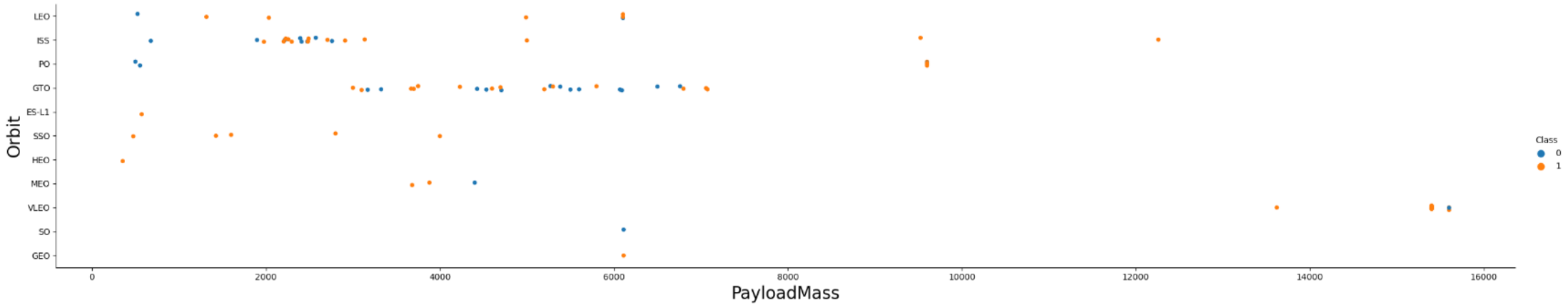
Orbit x Flight Number



The plot above shows that the success rate improved over time, that pattern it's seen in every orbit.

The orbit VLEO shows a positive success rate and an increase of the frequency, it appears to be a good business opportunity.

Orbit x Payload Mass



The plot above shows that there is no correlation between the variables, the better example is the orbit GTO, the orbit with most payloads.

The orbits SO and GEO have just one launch each, so there is almost no data to evaluate the success rate.

Yearly Launch Success Rate

Success rate started increasing in 2013 and kept until 2020;

The plausible analogy is that the first three years were a period of adjusts and improvement of technology.

CONCLUSION

- Different data sources were analyzed, refining conclusions along the process;
- The best launch site is KSC LC-39A;
- Launches above 7,000kg are less risky;
- Although most of mission outcomes are successful, successful landing outcomes seem to improve over time, according the evolution of processes and rockets;
- Decision Tree Classifier can be used to predict successful landings and increase profits.

APPENDIX

- All the plots, charts and graphs can be seen in the link